Biased Restaurant Reviews: Starbucks in NYC

Presenters: Hsuan-Hsueh Huang, Yuqi Yan, Jiangyue Mao

Motivation

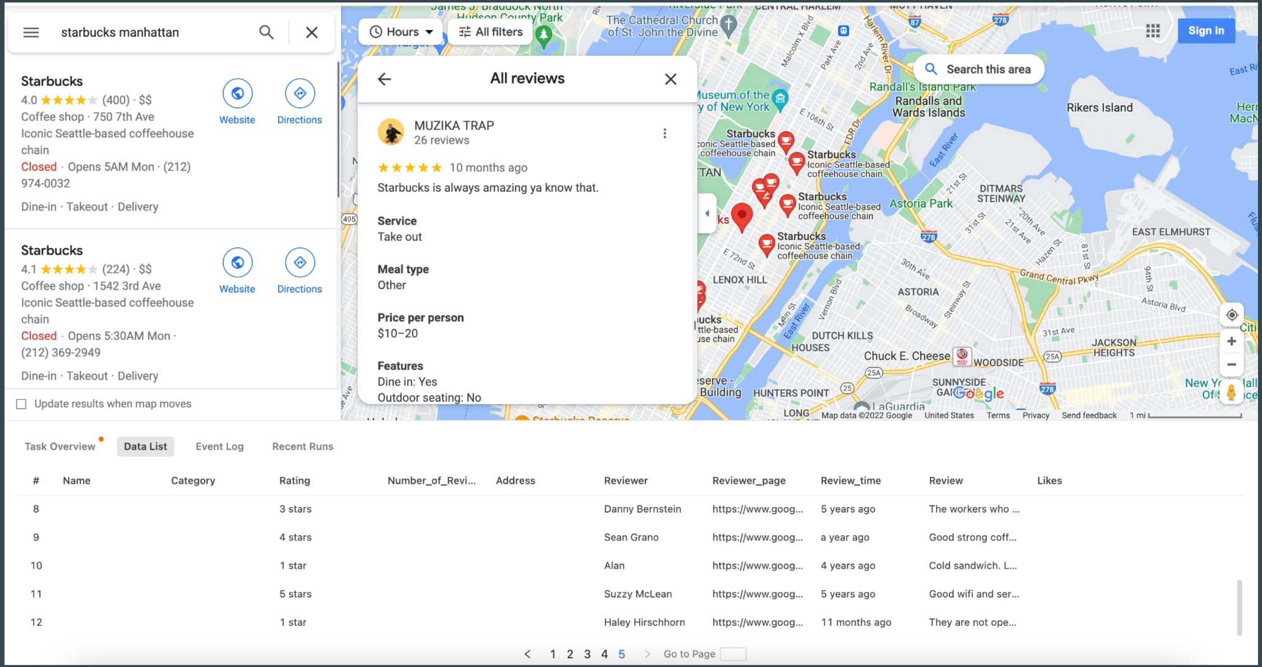
- Customers feedbacks are important for running the business
- The analysis result can be used to evaluate whether Starbucks stores are easy/hard to gain better ratings and reviews in a specific region
- Answer the following questions
    - What factors do customers care about Starbucks?
    - If those factors (extracted from reviews) varied by region?
        - If so, what makes them different?

Related works

- "Restaurant survival analysis with heterogeneous information."
  - Built a predictive model to find out whether a restaurant is still open 4 years later.
  - Used a similar dataset from Dianping (a Chinese app similar to Yelp) and created regional metrics for their analysis. Those metrics take into account the similar surrounding restaurants and also the density of restaurants nearby.
- "Social Influence Bias: A Randomized Experiment"
  - People tend to give ratings differently based on the treatments they received (e.g. no treatment, up-treated by positive social influence, etc.).
- "Is there a correlation between restaurant ratings and the income levels of a neighborhood?"
  - Business owners running a restaurant might spot an opportunity to gain customers by improving their customer service especially in low income areas.
  - Patrons living in low income neighborhoods would have the ammunition to demand better customer service in the restaurants they frequent.
  - City hall officials might also see this as a form of income segregation and work harder to ensure restaurants in low income areas have the same amenities and service levels as those in middle or high income areas.
- "Narrative framing of consumer sentiment in online restaurant reviews"
  - The more pricey the restaurant is, the longer the reviews it tend to receive.

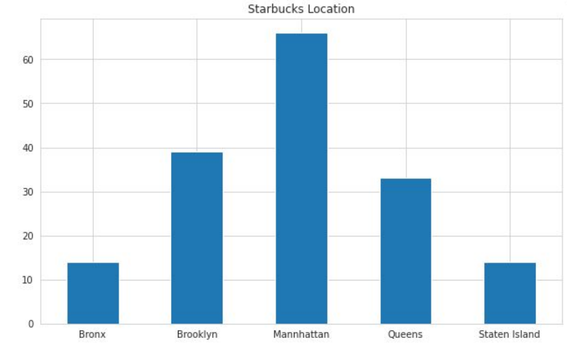
Review Data

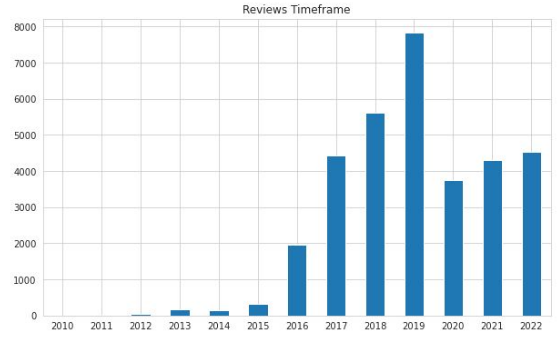
- Used a crawler to collect customers' reviews from Google map
- Attributes
    - Rating
    - Review
    - Review time
    - # Likes



Review Data

- Collected data from 5 borough in New York City
  - Totally 33156 reviews from 166 starbucks



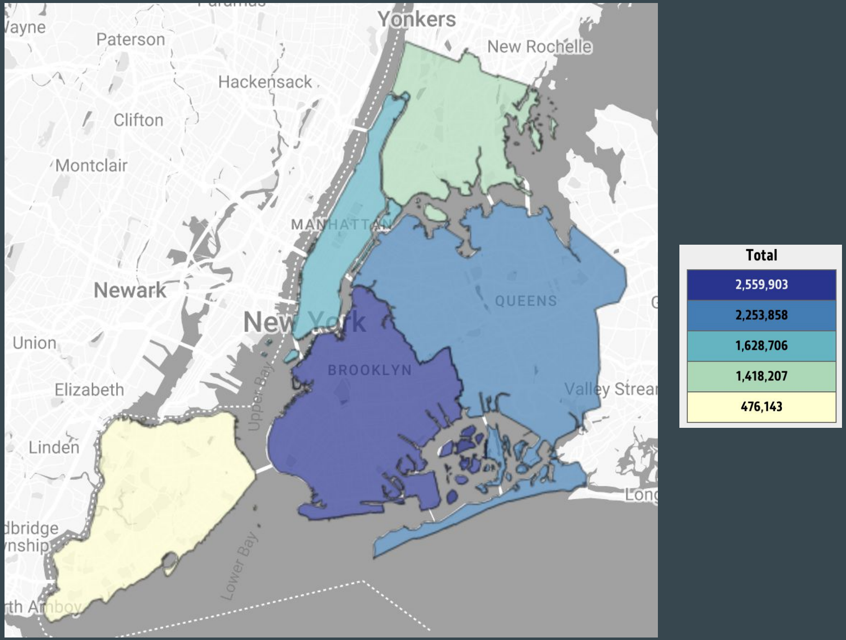


Regional Data

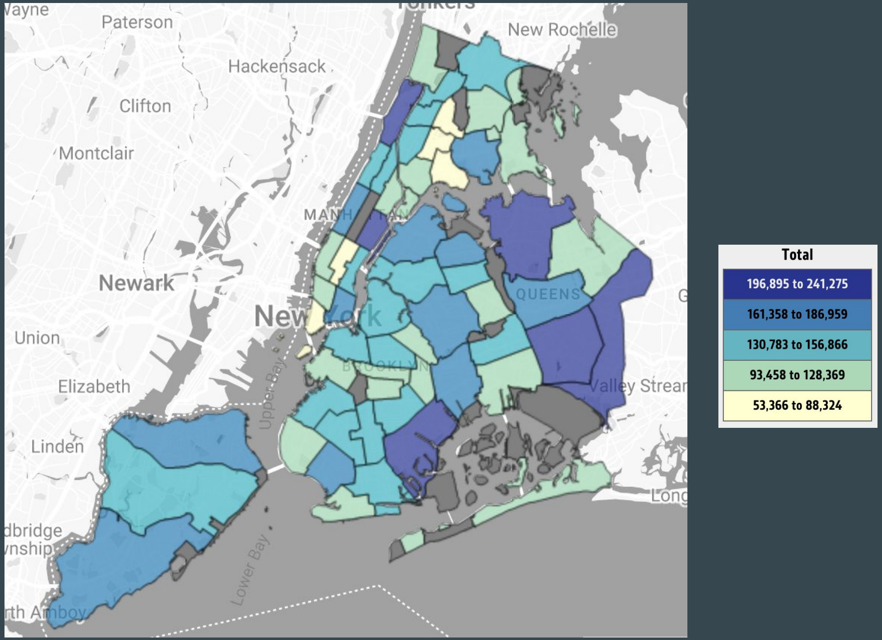
- Source: Citizen’s Committee for Children of New York
- Includes
  - Demographics
    - **Total population**, households & families, etc
  - Economics
    - **Household income**, poverty, etc
  - Education
    - **Postsecondary enrollment**, student performance metrics, etc
  - Community Safety
    - **Arrests**, Street & Sidewalk cleanliness, Reported Felonies, etc
  - Health & Mental Health
  - Housing & Homelessness

Regional Data - total population

Borough



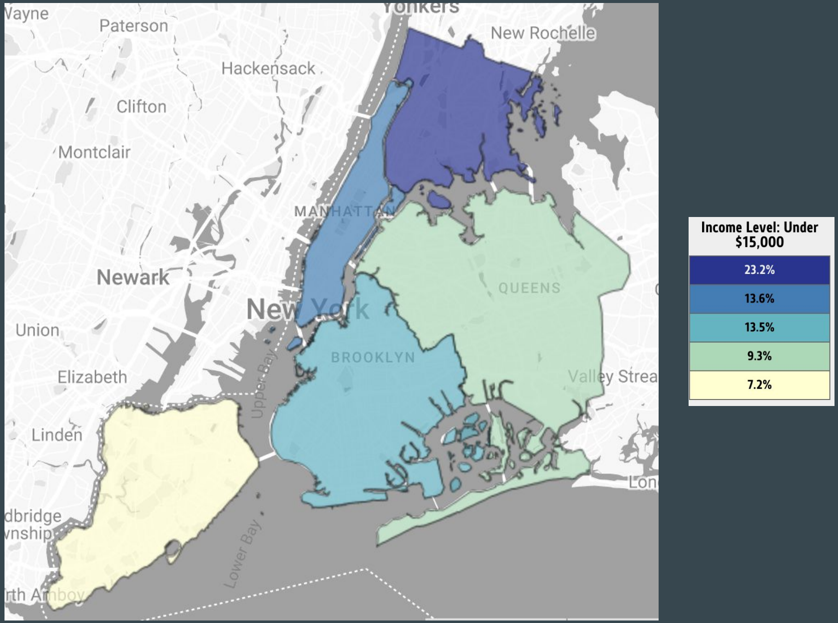
Community District



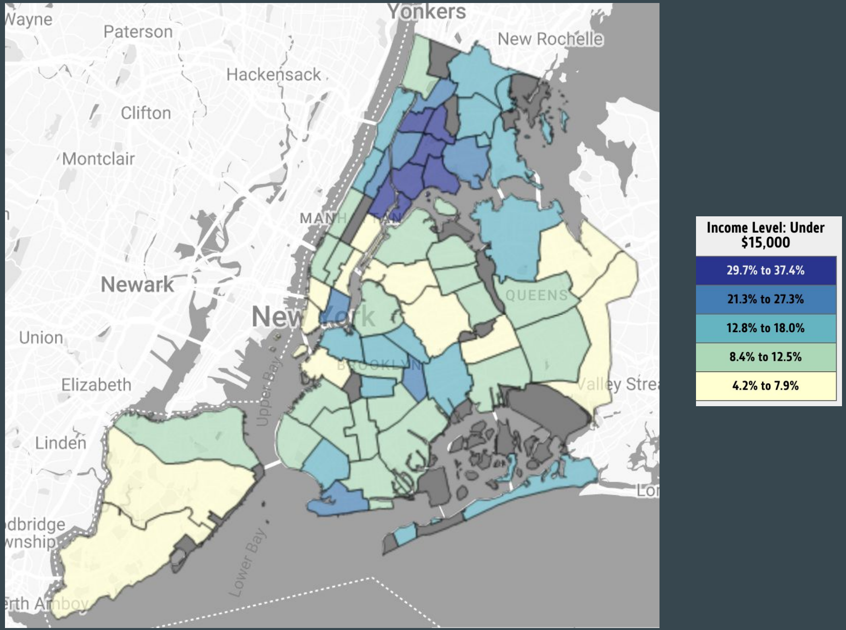
Source: https://data.cccnewyork.org/data/map/97/total-population#84/99/3/130/62/a/a

Regional Data - household income

Borough

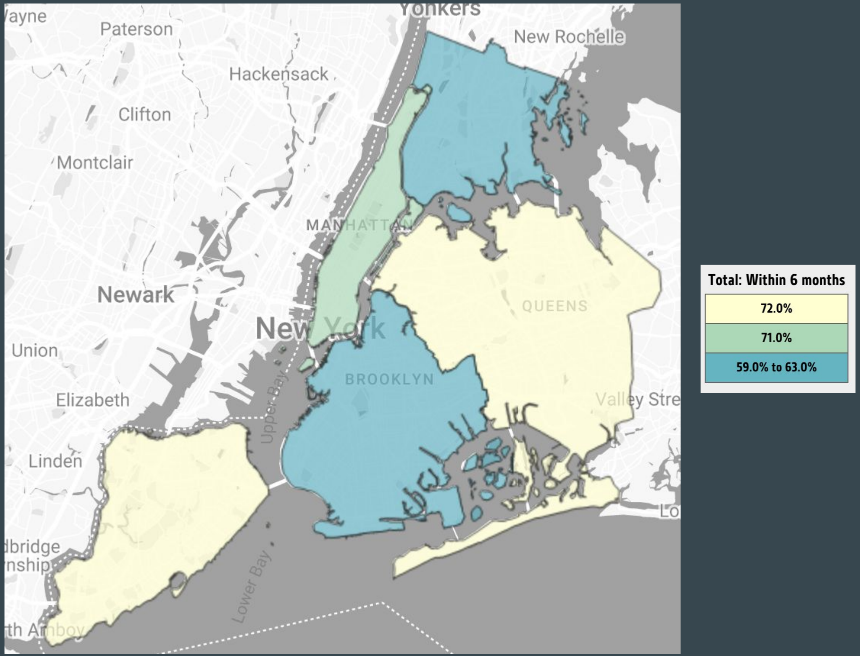


Community District

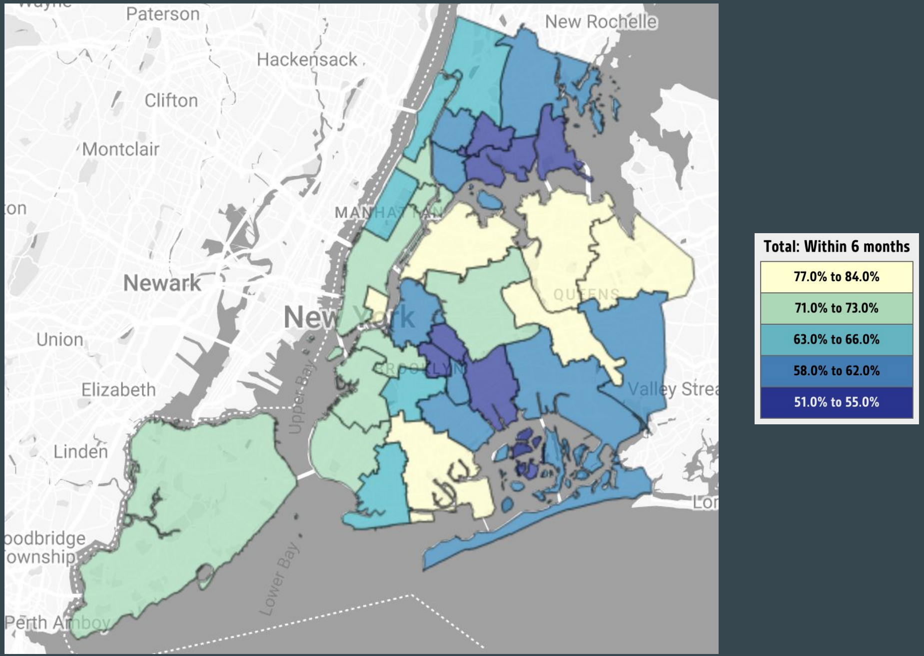


Regional Data - postsecondary enrollment

Borough

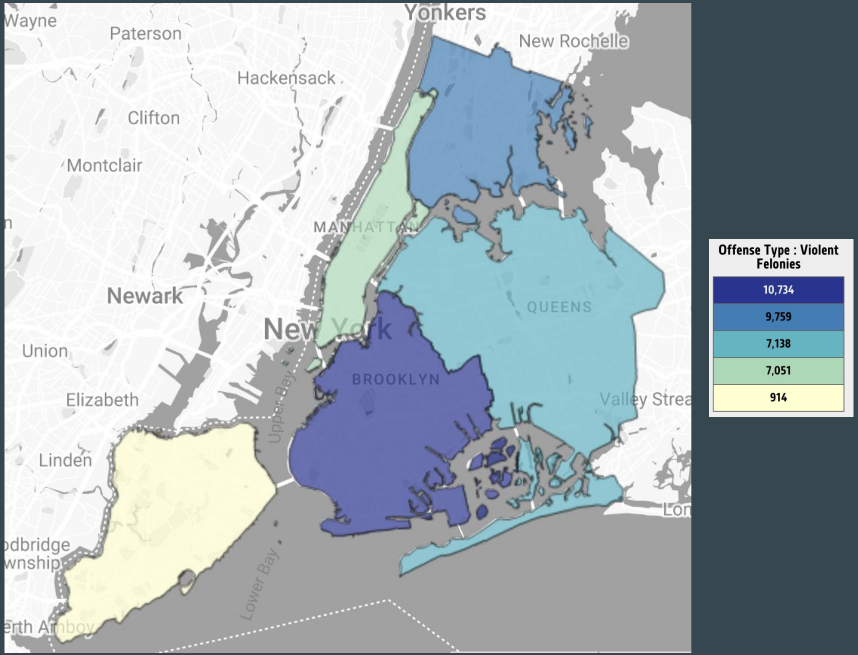


School District

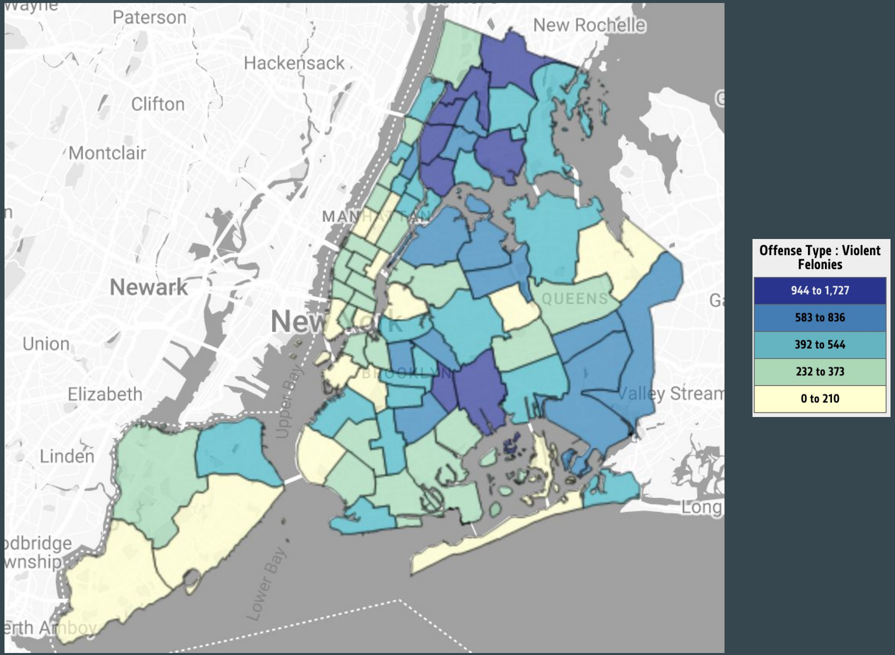


Regional Data - arrests

Borough



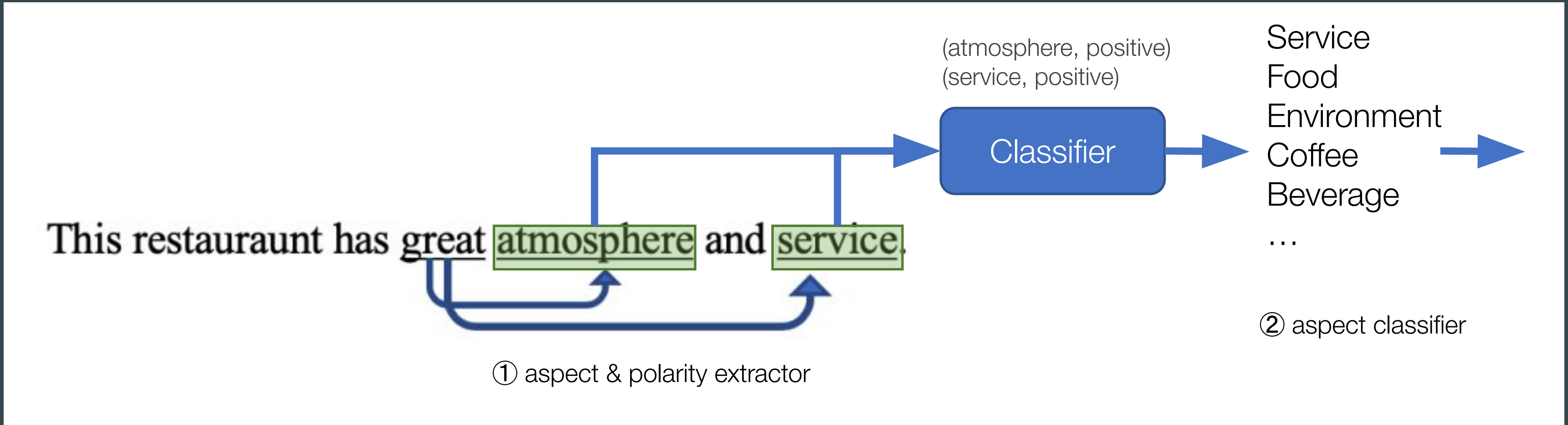
Police Precinct



Methods

To understand the customers' reviews and the related sentiment towards certain aspects of the Starbucks, we implemented two models:

1. Aspect Extractor: extract the opinion aspects & polarity from each review sentence
2. Aspect Classifier: classify the aspects for further analysis



Methods - Data Processing

- Before aspect extraction
  - Multilingual reviews
    - Google Map has integrated translator. We use English reviews and the translated results for other languages.
  - Emoji Removal (not supported in our aspect extractor)
  - Split into short paragraphs (limitation of the input size in the backbone network)
- Before aspect classification
  - Stemming
    - E.g. seat, seats -> seat
  - Vectorization with pre-trained language model (GloVe)

Methods - Aspect Extractor

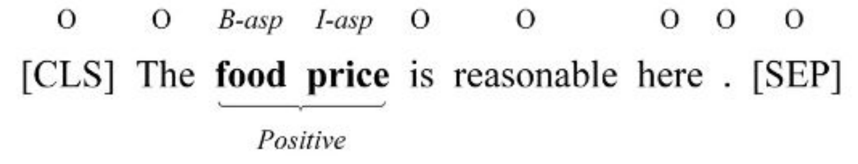
Table 1
Several samples from seven ABSA datasets. All the datasets are domain-specific.

No.	Sentence	Aspect	Polarity
1	Great laptop that offers many great **features**!	**features**	positive
2	The **seats** are uncomfortable if you are sitting against the wall on the wooden benches.	**seats**	negative
3	How do you settlers of catan for the **xbox**?	**xbox**	neural

"The **seats** are uncomfortable if ..." -> {w0, **w1**, w2, w3, w4, ...}

Label: {O, B, O, O, O, ...}

O: Non-aspect
B: beginning word of an aspect
I: non-beginning word

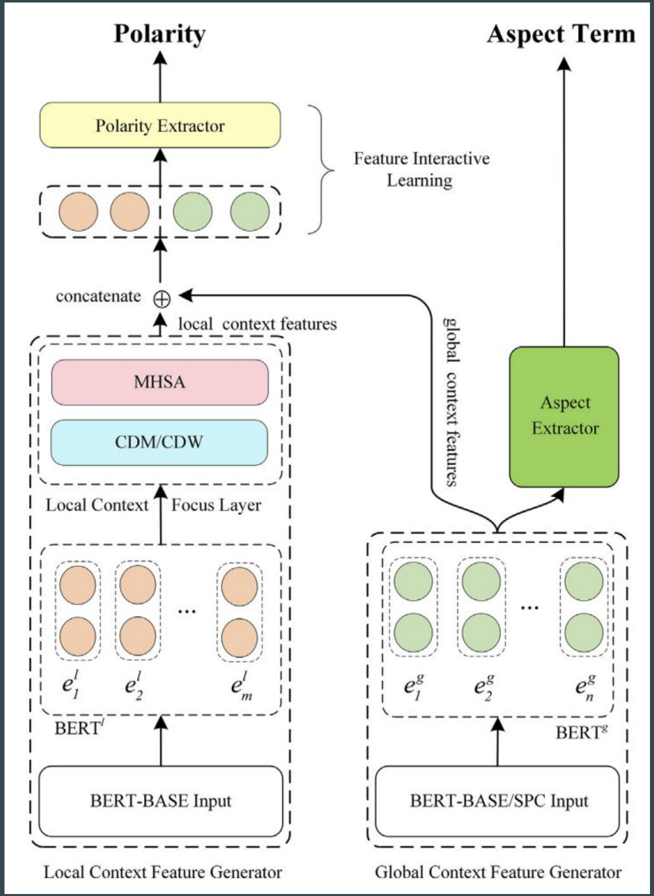


(a) The English input format of BERT-BASE.

Source: A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction
https://doi.org/10.1016/j.neucom.2020.08.001

Methods - Aspect Extractor

- ATEPC (**A**spect **T**erm **E**xtraction & Aspect **P**olarity **C**lassification)
  - Multi-learning network
  - Fusion: local context & global context
    - The identification of local context depends on semantic-relative distance (SRD). SRD describes how far a token is from a targeted aspect.
  - Domain-adapted BERT
    - Here we use the Restaurant reviews from SemEval-14 (alt.qcri.org/semeval2014/task4) as our training set. F-1 score: 0.805 in Rest14.
- This model suits our research question well for it's pre-trained restaurant-specific BERT
  - Assumption: Starbucks data distribution is similar to the general restaurant review data distribution (similar aspects)



Source: A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction
https://doi.org/10.1016/j.neucom.2020.08.001

Methods - Aspect Classifier

- Mini-batch K-means
  - A variant of the popular K-means clustering algorithm, with less computation cost
  - Instead of generating centroids for the whole dataset, Mini-batch K-means randomly picked a subset at each iteration
- We use unsupervised learning method for our research since we don't have labels for the aspects.

Algorithm 1 Mini-batch k-Means.

Given: k, mini-batch size b, iterations t, data set X
Initialize each $\mathbf{c} \in C$ with an $\mathbf{x}$ picked randomly from X
$\mathbf{v} \leftarrow 0$
for $i = 1$ to t **do**
    $M \leftarrow b$ examples picked randomly from X
    for $\mathbf{x} \in M$ **do**
        $\mathbf{d}[\mathbf{x}] \leftarrow f(C, \mathbf{x})$ // Cache the center nearest to $\mathbf{x}$
    end for
    for $\mathbf{x} \in M$ **do**
        $\mathbf{c} \leftarrow \mathbf{d}[\mathbf{x}]$ // Get cached center for this $\mathbf{x}$
        $\mathbf{v}[\mathbf{c}] \leftarrow \mathbf{v}[\mathbf{c}] + 1$ // Update per-center counts
        $\eta \leftarrow \frac{1}{\mathbf{v}[\mathbf{c}]}$ // Get per-center learning rate
        $\mathbf{c} \leftarrow (1 - \eta)\mathbf{c} + \eta\mathbf{x}$ // Take gradient step
    end for
end for

Source: Web-Scale K-Means Clustering
https://www.eecs.tufts.edu/~dsculley/papers/fastkmeans.pdf

Methods - Correlation Analysis

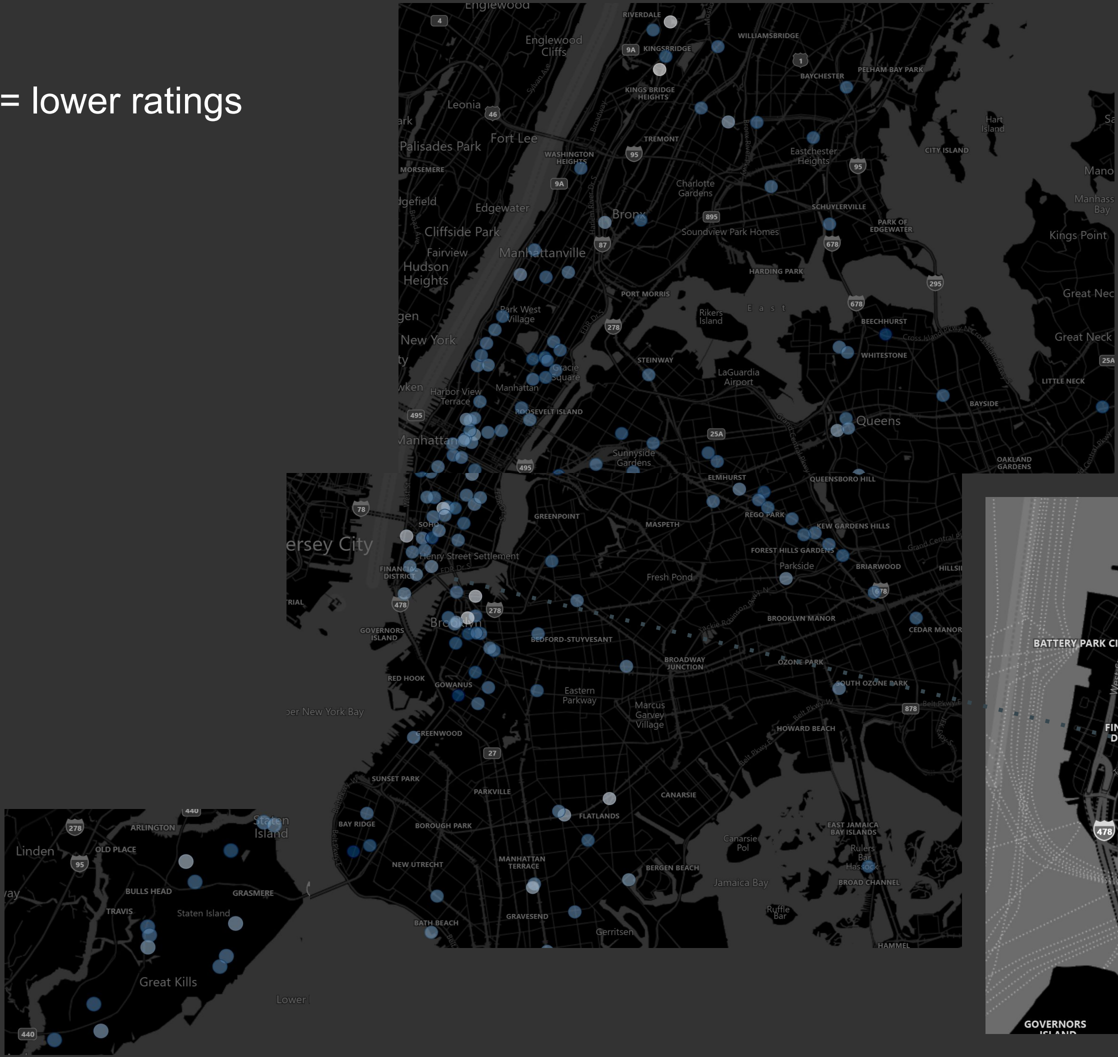
- Spearman’s Rank Correlation
  - measures the strength and direction of association between two ranked variables. i.e. how well the relationship between two variables could be represented using a monotonic function.

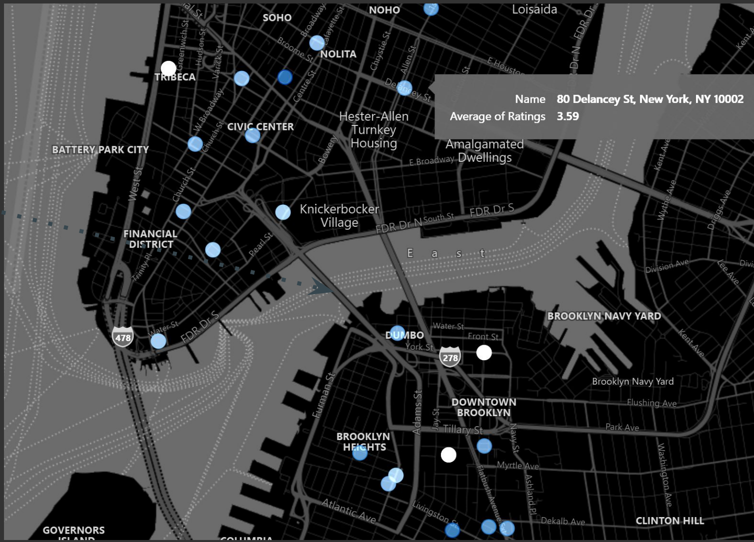
$$\rho = 1 - \frac{6\Sigma\, \mathrm{d}_i^2}{n(n^2 - 1)}$$

  - ρ = Spearman’s rank correlation coefficient
  - di = Difference between the two ranks of each observation
  - n = Number of observations
  - ρ takes value from -1 to +1. A value of +1 means a perfect association of rank.
- With pairwise Spearman’s rank correlation coefficient, we can find possible related factors (though not causal) of different sentiment towards different aspects of the restaurant.

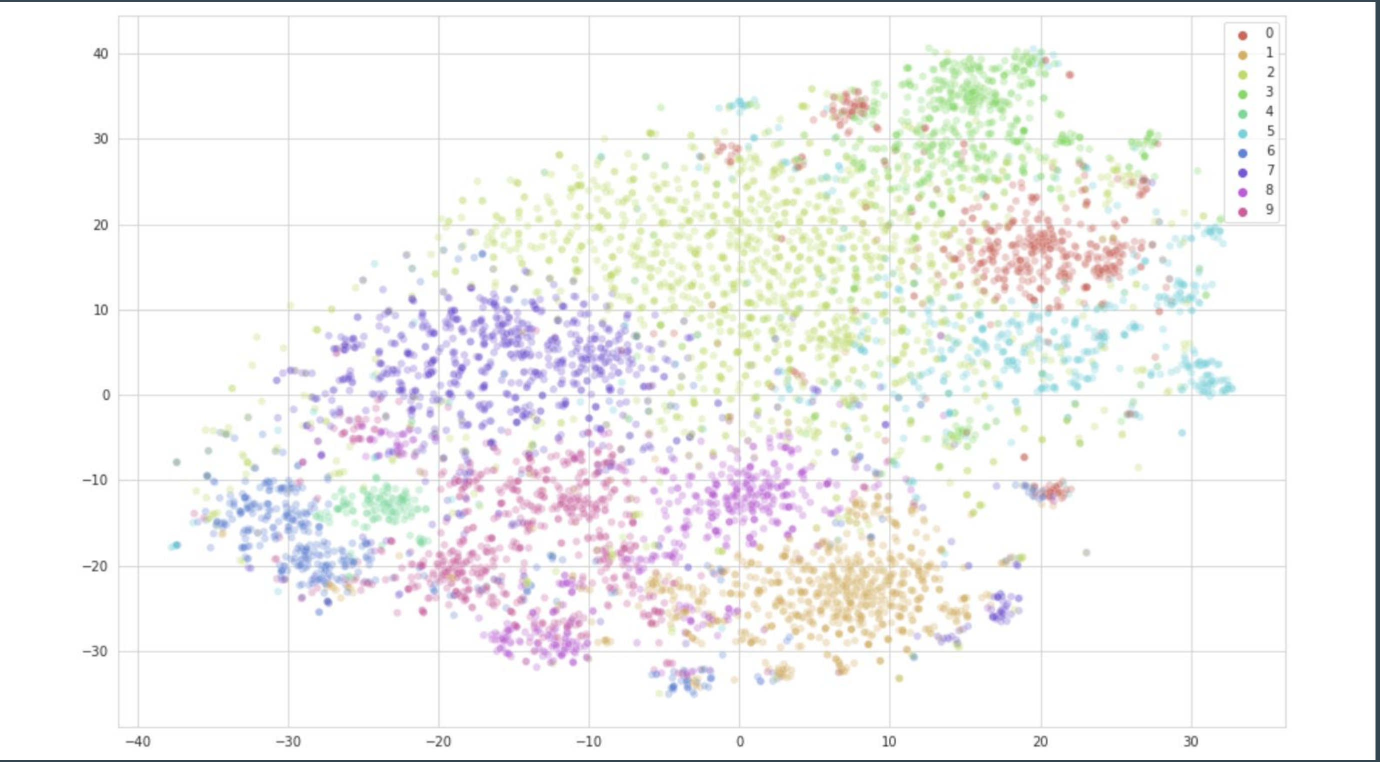
Results - Starbucks distribution

Lighter = lower ratings

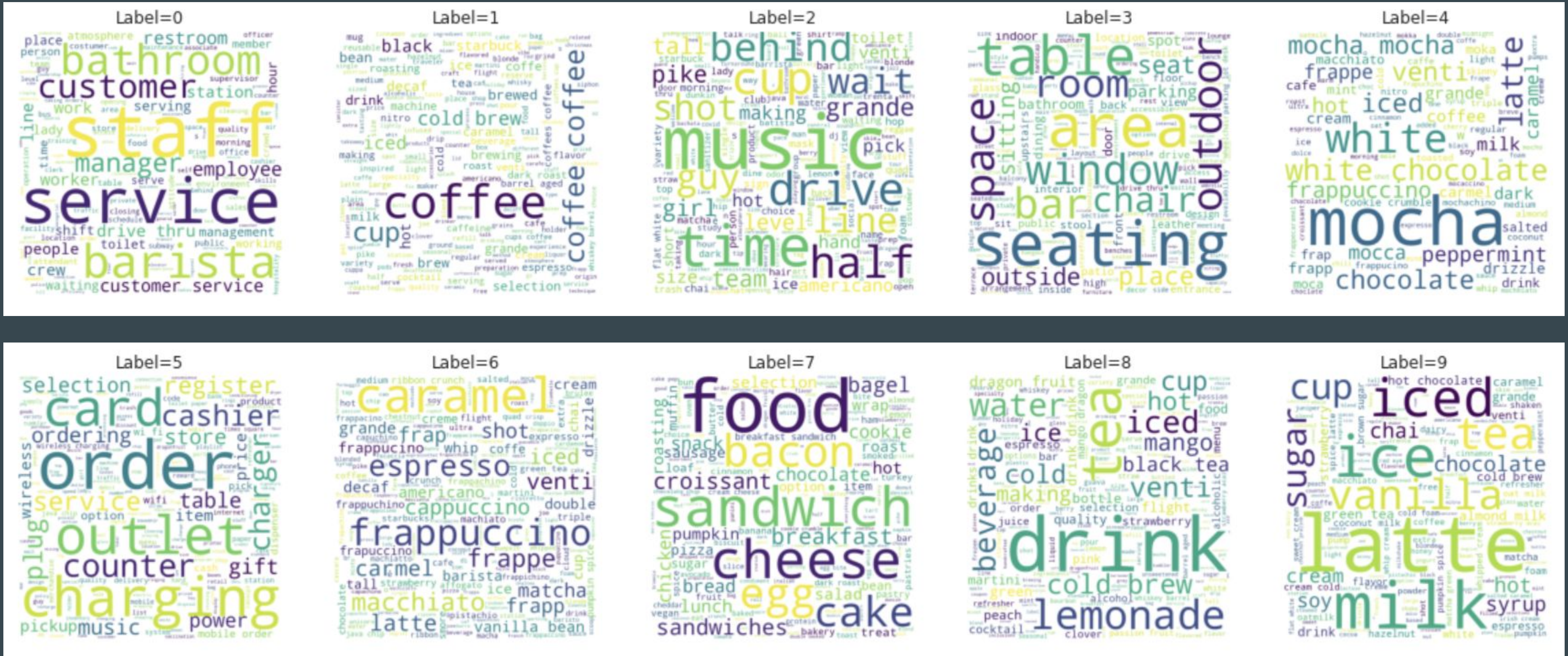




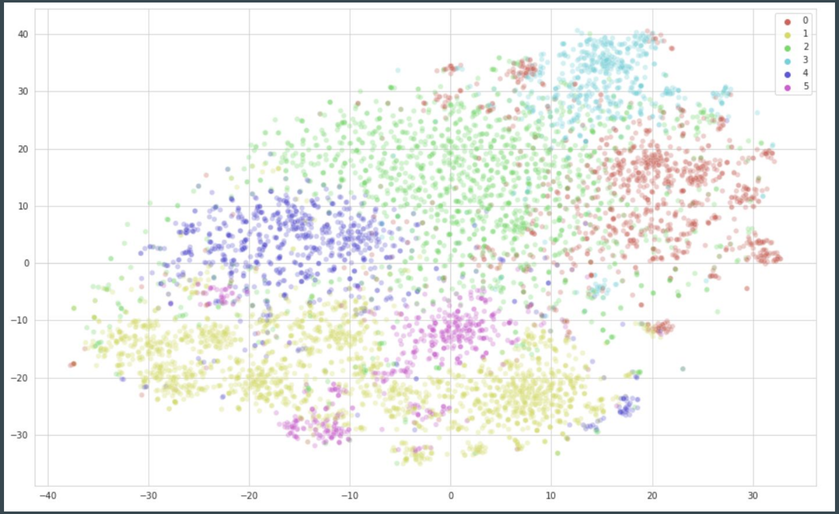
Lighter = fewer number of ratings



t-SNE: aspect embeddings
(7164 unique aspects)

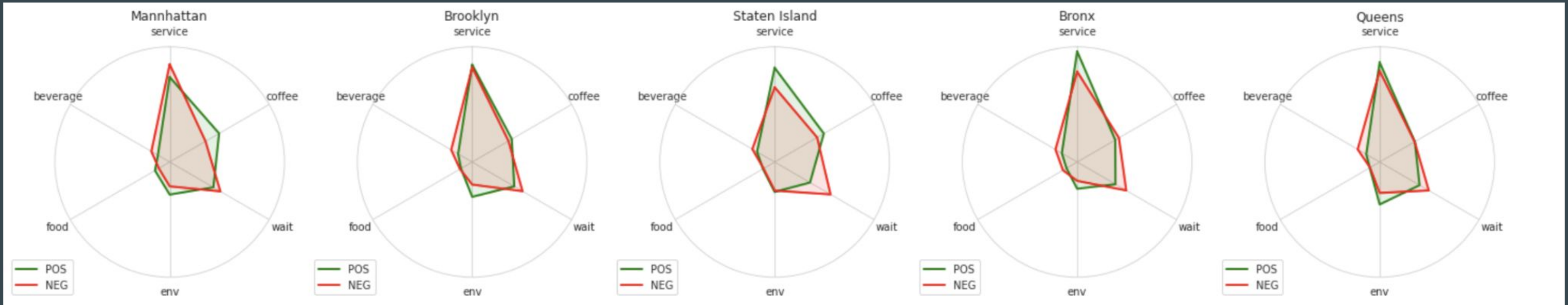


Aspect embeddings clusters
(K = 10)

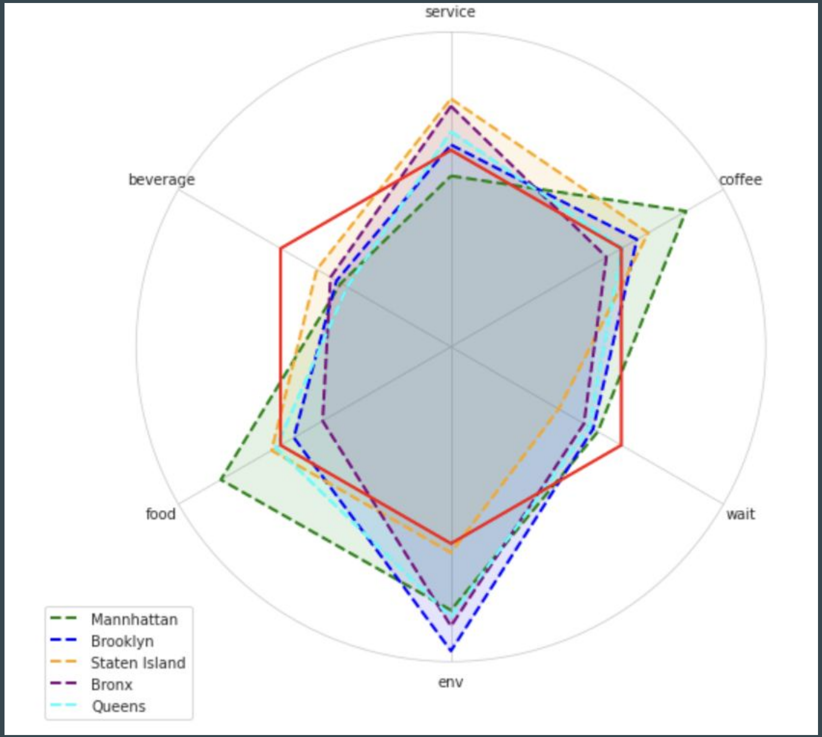


Merge similar clusters. 6 Clusters are selected.

["service", "coffee", "wait", "env", "food", "beverage"]

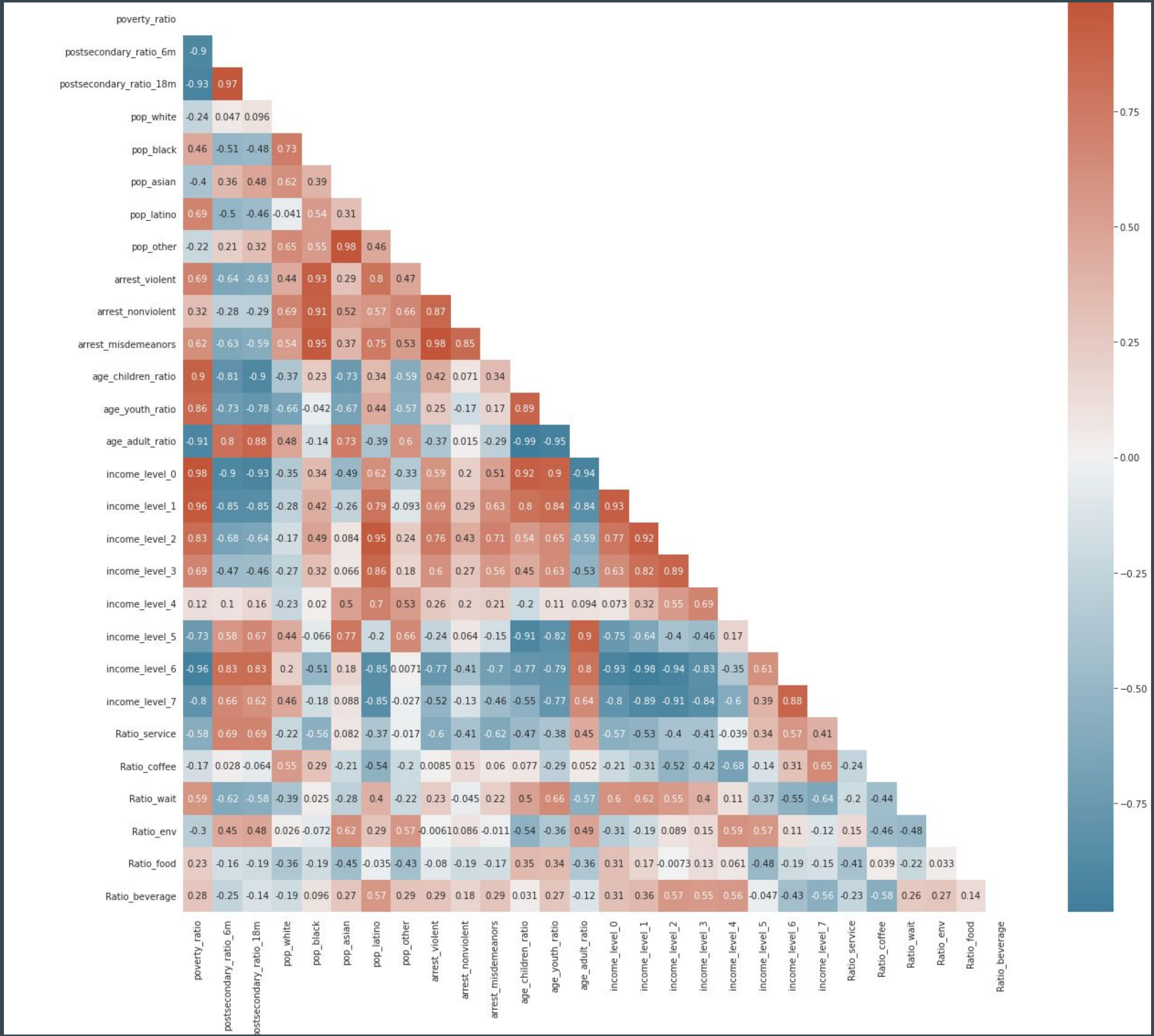


Pos & Neg reviews proportion

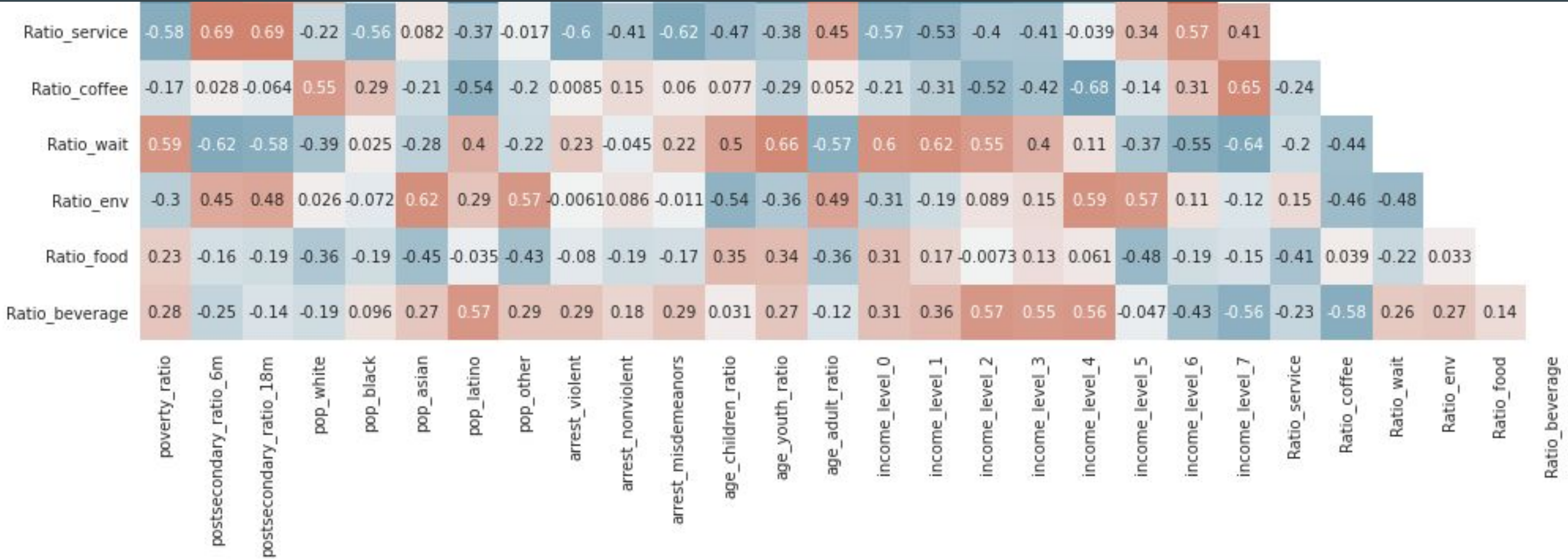


Pos / Neg proportion ratio

Correlation between demographic distributions and rating aspects



Income levels:
'$200,000 or more',
'$100,000 to $199,999',
'$75,000 to $99,999',
'$50,000 to $74,999',
'$35,000 to $49,999',
'$25,000 to $34,999',
'$15,000 to $24,999',
'Under $15,000'



Findings

- For NYC specifically, customers are more satisfied with the service and the environment of Starbucks in regions with lower income levels (and even higher crime rates) such as upper Manhattan and Bronx
    - Business owners might spot an opportunity to gain customers by improving their customer service, especially in low-income areas.
    - Patrons living in low income neighborhoods would have the ammunition to demand better customer service in the restaurants they frequent.
    - City hall officials might also see this as a form of income segregation and work harder to ensure restaurants in low income areas have the same amenities and service levels as those in middle or high income areas.

- For Starbucks’ reviews in Manhattan, customers also care about food and coffee when they give high ratings.
- Income levels and customers’ age distributions are the key factors influencing pos/neg ratio related to beverages and waiting time, and the ratio of beverages is also related to race distributions.

Limitations and future work

- Assumption: customers going to Starbucks have the same demographic distribution as the local citizens
  - Should use customers' demographic information instead. We made this assumption because the actual customers' information is unavailable. Real customers' information will give more accurate results in the correlation analysis.
- Aspects extractor is trained on Semeval-14 restaurant review dataset. The distribution may differ from our dataset.
  - The aspects extractor is important for our study, since we're interested in how people react differently to a restaurant (or, Starbucks in our case study)
  - Our dataset doesn't have aspect labels, therefore it's necessary to hand-label several reviews (ideally ~1k) to see how the aspects extractor works. The model can be improved by fine-tuning on the labeled dataset.

Limitations and future work

- Aspect classification is done by using an unsupervised learning method. We don’t have control over how the aspects are classified.
  - From the visualization result, this method makes sense and suits our research question. For future study in specific aspect categories, we may need to figure out what categories we should use, label the aspects and use supervised methods instead.
- Insufficient data points
  - Due to the limited availability of demographic information, we mainly focus on Starbucks near NYC. These findings may or may not be applied to other coffee shops or regions.